Building the Learning Analytics Curriculum: Should we Teach (a Code of) Ethics?

Conference or Workshop Item

Building the Learning Analytics Curriculum: Should we Teach (a Code of) Ethics?

Paul Prinsloo
Department of Business Management
University of South Africa
prinsp@unisa.ac.za

Sharon Slade
Faculty of Business and Law
The Open University
sharon.slade@open.ac.uk

Abstract

This brief chapter explores the feasibility of teaching (a code of) ethics against a background which examines our views around data scientists, data analysis, data and, in particular, student data. It touches upon different approaches to ethics and asks whether teaching ethics would make any difference.

1. The need for an ethical approach to the collection, analysis and use of student data

There are many reasons to consider not only the ethical implications in the collection, analysis and use of student data, and also how such challenges might be addressed ([18], [19], [23]). Trends in international higher education and ideological positions pertaining to the purpose of higher education will shape the scope and appropriateness of our responses to ethical issues in learning analytics [16].

Foundational to our approach and proposal is the consideration of learning analytics as a non-neutral structuring device, informed by our current beliefs about what counts as knowledge and learning. Further, approaches to coding, analysis and interpretation are inevitably colored by assumptions about gender, race, class, capital and literacy, and so are both in service of and perpetuate existing power relations. We suggest that (a code of) ethics in learning analytics should include how we see data, what we see as data, the purposes and processes for collecting data, our data analyses (who does this, how it is verified, how/why the output is shared etc), and whether higher education institutions have the resources to ethically respond to identified needs or gaps. Codes of ethics or curricula focused on the ethical issues in data usage should do more than simply provide information but should also aim to change behavior or at least, prompt reconsideration of entrenched positions and values. Designing these codes and curricula requires consideration of different approaches to ethics as well as the conditions that will enable these codes and curricula to shape practice.

2. Examining beliefs and assumptions

Efforts to design codes of ethics or indeed curricula that recognize ethical issues can often underestimate the extent to which curriculum acts as contested space in which different beliefs and assumptions play out [17]. Central to our discussion is the need to consider beliefs and assumptions about the role and identity of data scientists, data analysis, and data itself.

Data scientists are portrayed as having the ‘sexiest job of the 21st century’ [4]; as being the new ‘rock stars’ [20] and as high priests of algorithms [6]. In somewhat stark contrast to this social imaginary, Walker [25] paints data scientists as fallible humans with biases, suggesting that:

> Humans … interpret meaning from data in different ways. Data scientists can be shown the same sets of data and reasonably come to different conclusions. Naked and hidden biases in selecting, collecting, structuring and analyzing data present serious risks. How we decide to slice and dice data and what elements to emphasize or ignore influences the types and quality of measurements [25, p. 11]

A recent report by Harris, Murphy and Vaisman [10] also points to the huge variety of skill sets and backgrounds of those who might classify themselves as 'data scientists'. Not only do they come from a range of academic/disciplinary backgrounds such as, inter alia, mathematics, statistics, computer science, researchers, business people and software developers; their expertise is often focused on one particular domain of knowledge or practice such as data mining, big data, analyzing structured and unstructured data, statistical modeling, and/or programming. In the light of the increasing velocity, variety and volumes of data [13] it seems that future data analyses will require multiple skills and backgrounds found in teams, rather than individuals.

In addition to the social imaginary pertaining the power of data scientists, we should also consider data analysis practices. Data analysis has been described as an "art" [11] and as "black art" [8]. Given the increasing complexities of software and methods used to engage with and analyze data, and that many analyses are described in very technical language and terminology, there is an impression that data analysis provides access to 'hidden' knowledge, not normally accessible to mere mortals. And this then means that many users of those analyses require an 'interlocutor' to provide them with an understanding of the data and findings.

The last set of beliefs and assumptions that requires our attention is around data itself. These include that data are neutral and represent the 'Truth' – you can't argue with data. We talk about data as "raw", "cooked", "corrupted", "cleaned", "scraped", "mined" and "processed" [9]. We assume that this data speaks for itself [14], that is, that there is no further need to interpret or extract any underlying meaning. There can be a tendency to assume that a given dataset represents the whole picture (especially in big data), and that knowing 'what' is happening erases the need to know 'why' it may be happening. Many researchers and analysts believe that big(ger) data are better data, often not heeding the warning by Silver [22] that bigger data sets make it harder to distinguish between the signal and the noise.

Central to any consideration of the ethical implications in the collection, analysis and use of data must be the principled position that data are *not* neutral, raw, objective and pre-analytic, but framed "technically, economically, ethically, temporally, spatially and philosophically. Data do not exist independently of the ideas, instruments, practices, contexts and knowledges used to generate, process and analyse them" [13, p. 2]. Often when presented at forums or in reports, data are presented as indisputable, as fact. As researchers, we will be aware that when a given theory is proven false, it is no longer accepted as a fact. The same position is rarely applied to data…

2.1. What are the implications for learning analytics as ethical practice when data are framed and framing?

With the above as points of departure, we should consider that relationships between data, information, knowledge, evidence and wisdom are much more complex and contested than we may be comfortable with (e.g., [13]). If we were to accept data as neutral, objective and value free, we might also be expected to accept that there is a 'right' way to analyze data, and that scientists are to be respected and never questioned. In such a case a code of ethics or a curriculum which deals with ethics in data would look very different from a code or curriculum that accepts data as, per se, framed and framing.

So, we understand the need for a code of ethics and a curriculum that addresses the ethical implications in the collection, analysis and use of data in the light of our belief that data *are* political in nature – data are loaded, shaped and limited with the values, interests and assumptions of those who collect, frame and use them [21]. A code of ethics and a curriculum dealing with the ethical issues in the collection, analysis and use of data must consider the danger of apophenia, that of "seeing patterns where none actually exist, simply because enormous quantities of data can offer connections that radiate in all directions" [2, p. 668]. (Also see [8]).

2.2. What are the implications for learning analytics when the student data we've collected and analyzed are not the whole picture?

There is ample evidence that suggests that the collection, analysis and use of student data is seen as a 'revolution' and as the solution to many, if not most, of the challenges in current higher education. Student data are seen as the 'new black" [26], as oil, as a resource to be mined. We believe that our data dossiers, and increasingly our access to students' behavioral digital data allows us to compile *complete* pictures regarding students' aspirations, their potential and their life journeys. There is, furthermore, an inherent assumption that student data are owned by the higher education institution which collects it, that students don't need access, or to know what we collect, the reasons for the collection, how we analyse the data, how long we keep the data, who has access to the data, and who we share the data with. And equally worrying, the question of whose interests are really at stake is rarely raised.

While we proclaim ‘student-centeredness’ and putting ‘students first’, students are neither informed, nor involved in the analysis of data, the provision of context, provision of additional data or able to inform institutions what data *they* would like to have access to in order to make more informed decisions. Students therefore may have reason to question whose values and interests are mainly served in the collection, analysis and use of *their* data.

2.3. Whose values? Whose ethics?

Prinsloo [16] flags the tensions and pull within higher education between *liberal* values (serving the public good, a focus on increasing equality, as key to the delivery of national goals), *neoliberal* values (introducing the commodification of the curriculum, a focus on students as customers and increasing educational administration) and a *socio-critical* orientation (exploring the inherent epistemological power in curricula and raising the notion of the university as an elitist space). Though it falls outside of the scope to discuss the implications of each of these three approaches (liberal, neoliberal and socio-critical) to the role of higher education in the 21st century, each of these orientations will have an impact on the collection, analysis and use of student data (also see [5]).

In addition to categorizing the role of higher education, we should also consider different approaches to ethical frameworks, such as deontological and teleological approaches. A deontological approach to ethics (whether as code or curriculum) typically forms the basis for legal and regulatory frameworks, Terms and Conditions, contractual agreements and simple opt-in/opt-out approaches to sharing of personal data. A deontological approach works best in stable environments. In contrast, we can also consider a teleological approach which emphasises potential for harm, individuals’ agency in making informed, nuanced decisions regarding the collection, analysis and use of their data and recourse to action in cases of harm or breach of privacy (see [1], [18], [19], [27]).

Velasquez, Andre, Shanks, and Meyer [24] suggest a combination of approaches that may include the following:

- A **utilitarian** approach (action that “provides the greatest balance of good over evil”);
- A **rights approach** (referring to basic universal rights, such as the right to privacy, not to be injured, etc);
- A **fairness** or **justice** approach;
- A **common-good** approach (the welfare of the individual is linked to the welfare of the community); and
- A **virtue** approach (based on the aspiration towards certain shared ideals).

It is clear then that the scope and content of a code of ethics and/or a curriculum addressing the ethical issues in the collection, analysis and use of student data is more complex than may at first be suspected. And, in any case, there is a concern that simply establishing codes of ethics or curricula which more explicitly deal with ethical issues makes little difference…

3. (In)conclusion: Codes of ethics – who cares anyway?

It falls outside the scope of this short article to consider the different opinions regarding the impact of codes of ethics or curricula that teach ethics (see, e.g., [7]). Numerous codes of practice and policies relating to ethical uses of learning analytics have been produced (e.g., [12], [15]) and it is perhaps too early to say whether simply developing such frameworks changes practice.

We argue that “Ethics are the mirror in which we evaluate ourselves and hold ourselves *accountable*” and that holding actors and humans accountable still works “better than every single other system ever tried” ([3], emphasis added). And the question is really not “should we teach (a code of) ethics as part of a learning analytics curriculum?”, but under what conditions might this actually make a difference?

4. References

[1] Botes, A. (2000). A comparison between the ethics of justice and the ethics of care. Journal of Advanced Nursing, 32:1071–1075. DOI: 10.1046/j.1365-2648.2000.01576.x

[2] boyd, D., & Crawford, K. (2012). Critical questions for big data. Information, Communication & Society, 15(5): 662-679. http://dx.doi.org/10.1080/1369118X.2012.678878

[3] Brin, D. (2016). Preparing for our posthuman future of Artificial Intelligence. Retrieved from https://omni.media/preparing-for-our-posthuman-future-of-artificial-intelligence

[4] Davenport, T.H., & Patil, D.J. (2012). Data Scientist: The sexiest job of the 21st century, Harvard Business Review, Oct 2012). Retrieved from https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

[5] de Oliveira Andreotti, V., Stein, S., Pashby, K., & Nicolson, M. (2016). Social cartographies as performative devices in research on higher education. Higher Education Research & Development, 35(1): 1-16.

[6] Dwoskin, E. (2014). Big Data's high priests of algorithms. The Wall Street Journal, Aug 8, 2014. Retrieved from https://www.wsj.com/articles/academic-researchers-find-lucrative-work-as-big-data-scientists-1407543088

[7] Fisman, R. & Galinsky, A. (2012). Can you train business school students to be ethical? Slate, Sept 4. Retrieved from http://www.slate.com/articles/business/the_dismal_science/2012/09/business_school_and_ethics_can_we_train_mbas_to_do_the_right_thing_.html

[8] Floridi, L. (2012). Big data and their epistemological challenge. Philosophy and Technology, 25(4): 435–437.

[9] Gitelman, L., & Jackson, V. (2013). Introduction, in Lisa Gitelman (ed.), "Raw data" is an oxymoron, (pp.1-14). London, UK: The MIT Press

[10] Harris, H.D., Murphy, S.P., & Vaisman, M. (2013). Analyzing the analyzers. An introspective survey of data scientists and their work. O'Reilly. Retrieved http://www.oreilly.com/data/free/files/analyzing-the-analyzers.pdf

[11] Ibrahim, M. (2015). The art of data analysis. Journal of Allied Health Sciences, 1(1): 98-104. Retrieved from https://www.researchgate.net/publication/283269432_The_art_of_Data_Analysis

[12] Jisc (2014). Code of practice for learning analytics: A literature review of the ethical and legal issues. Jisc, November 5. Retrieved from http://repository.jisc.ac.uk/5661/1/Learning_Analytics_A-_Literature_Review.pdf

[13] Kitchin, R. (2014). The data revolution. Big data, open data, data infrastructures and their consequences. London, UK: SAGE. http://dx.doi. org/10.4135/9781473909472

[14] Mayer-Schönberger, V., & Cukier, K. (2013). Big data. A revolution that will transform how we live, work, and think. New York, N.Y.: Houghton Mifflin Harcourt.

[15] Open University. (2014). Policy on ethical use of student data for learning analytics. Retrieved from http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/ethical-use-of-student-data-policy.pdf

[16] Prinsloo, P. (2016a). A social cartography of student data: Moving beyond #StudentsAsDataObjects. Retrieved from https://www.slideshare.net/prinsp/a-social-cartography-of-student-data-moving-beyond-studentsasdataobjects?

[17] Prinsloo, P. (2016b). Curricula as contested and contesting spaces: Geographies of identity, resistance and desire. Presentation at Transforming the Curriculum: South African Imperatives and 21st Century Possibilities, University of Pretoria, 28 January. Retrieved from https://www.researchgate.net/publication/292502252_Curricula_as_contested_and_contesting_spaces_Geographies_of_identity_resistance_and_desire

[18] Prinsloo, P., & Slade, S. (2016a). Student vulnerability, agency, and learning analytics: An exploration. Journal of Learning Analytics, 3(1): 159–182.

[19] Prinsloo, P., & Slade, S. (2016b). Here be dragons: Mapping student responsibility in learning analytics. In M. Anderson & C. Gavan (Eds.), Developing effective educational experiences through learning analytics (p. 170–188). Hershey, PA: IGI Global.

[20] Sadkowsky, T. (2014, July 2). Data Scientists: the new rock stars of the tech world, Technopedia. Retrieved from: https://www.techopedia.com/2/28526/it-business/it-careers/data-scientists-the-new-rock-stars-of-the-tech-world

[21] Selwyn, N. (2014). Data entry: towards the critical study of digital data and education. Learning, Media and Technology, 40(1): 64-82, http://dx.doi.org/10.1080/17439884.2014.921628

[22] Silver, N. (2012). The signal and the noise: Why most predictions fail – but some don't. New York, NY: Routledge.

[23] Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist, 57(1): 1509–1528.

[24] Velasquez, M., Andre, C., Shanks, T.S.J., & Meyer, M.J. (2015, August 1). Thinking ethically. Retrieved from https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/thinking-ethically/

[25] Walker, M. A. (2015). The professionalisation of data science. International Journal of Data Science, 1(1): 7-16.

[26] Watters, A. (2013). Student data is the new oil: MOOCs, metaphor, and money. [Blog post]. Retrieved from http://hackeducation.com/2013/10/17/student-data-is-the-new-oil

[27] Willis, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective. Educational Technology Research and Development, 64(5): 881-901.